Module 7 Lab Activity: Statistical Inference in SLR
PSY 652 Research Methods
Oct 14, 2020

1. Create a new R notebook from your project file and name it "slr_inference"
2. Create a new R chunk with a first level header: "Load Libraries"
   a. load the packages in this R chunk: tidyverse and psych
3. Create a new R chunk with the first level header: "Simulate the population"
   a. In your R chunk Copy and paste the following code:

```
# generate some data
set.seed(1870)
height <- rnorm(n=50000, m=70, sd=3.00) # generate a population of 50,000 (Avg Height of Men in the United States in inches)
my_pop <- data.frame(height)
```

   b. Run the chunk. You should now have a dataset with 50,000 observations of 1 variable available in your environment. This is your simulated population.
4. Insert a new R chunk with the first level header: "Describe data"
   a. In the R chunk, use the describe() function to summarize your dataset object (meanpop)
5. Create a first level header that says: "Create a histogram of the population"
   a. Use ggplot to create a histogram of height. Set the histogram binwidth to 0.30.
   b. Additionally, add the following:
      i. A title that says: "Distribution of height for U.S. Men
      ii. A subtitle that says: "Simulation of a population"
      iii. Label the x-axis as: "Height"

   Hint:

```
ggplot(data, aes(x = var)) +
      geom_histogram(binwidth = .30) +
      labs(title = "Title here", subtitle = "subtitle here", x = "x-axis here")
```

6. Create a first level header: "Randomly select 3 samples of 100 from the population"
   a. In an R chunk use the sample_n() function to randomly pull 100 participants from your population dataset *with replacement*. Save these each into their own object called *s1*, *s2*, and *s3*.
7. Create a first level header: "Describe each of the 3 samples"
   a. In an R chunk, use the describe() function three times to describe each of the sample datasets you just made (s1, s2, & s3)
      i. Compare these to the original population summary. Do the numbers look relatively similar? Although not exact, the mean value should be close to the true population mean.

8. Create a first level header: Inference for SLR
9. Create a second level header: "Simulate the population SLR results
    a. In an R chunk, copy and paste the following code:

```r
# Generate some data
set.seed(83587)
weight <- rnorm (mean = 197.0, sd = 1.2, n = 50000) #Create weight var Average weight
of men in the U.S. in pounds

b0 <- 0 # Pop intercept
b1 <- 1 # Pop slope
sigma <- 1.0 # Pop SD

e <- rnorm(mean = 0, sd = sigma, n = 50000) #generate error

height <- b0 + b1*weight + e #create height

my_regpop <- data.frame(weight, height) #Dataset
```

    a. Run the chunk. You should now have a dataset with 50,000 observations of 2
       variable available in your environment. This is your simulated population for
       height and weight.
10. Create a first level header: "Get a scatterplot of the population"
    a. In an R chunk, use ggplot to create a scatteplot with weight on the x-axis and
       height on the y-axis. Add a regression line (geom_smooth()) to the plot. Give the
       plot a snazzy title and label your x and y axes.

Hint:
```r
ggplot(data, aes(x = var1, y = var2)) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs(title = "Snazzy title", x = "name1", y = "name2")
```

11. Create a second level header: "Fit a SLR model to the population"
    a. In an R chunk, regress height on weight. (Weight is your predictor, height is the
       outcome variable). Save it to a new object called "popmod".
    b. In the same R chunk, display a summary of the model using summary()
12. Create a second level header: "Randomly select 3 samples of 100 from the population"
    a. In an R chunk use the sample_n() function to randomly pull 100 participants from
       your population dataset *with replacement*. Save these each into their own object
       called *ss1*, *ss2*, and *ss3*.
13. Create a second level header called: "Run a SLR on each sample"
    a. Run a regression analysis on each newly created object (ss1, ss2, & ss3) with
       height regressed on weight (The same equation you did for your population
       dataset). Save them into their own unique objects called r1, r2, r3.

      b. In the same R chunk, run summary()on each object (r1, r2, & r3). You will have 3 summary() statement calls.
           i. View the output from each of the regression outputs. Are they similar? They are expected to be slightly different but will be around the population parameters.

14. Once you've completed all of these steps, Restart R and Run All Chunks, and then preview your notebook. Save your notebook as both a .Rmd and an html file and exit RStudio.

15. Upload both the .Rmd and html version of your notebook to the assignment called "Module 7 Lab Activity" on the course Canvas page.